AI DIGEST

Anthropic mathematician claims Riemann proof with Claude

Nasdaq erased its 2026 gains, OpenAI guards reasoning from distillation, and Signal blocks AI agents on phones.

Saturday, 21 November 2026

IN THIS ISSUE

1. Alpöge and Claude announced a proof of the generalized Riemann hypothesis
2. On the same day: a human proof of GRH from Masullo, and people already look for holes in it
3. Market: Anthropic -9% on day two, Nvidia -14%, Nasdaq lost its 2026 growth
4. OpenAI on distillation: attack success on hidden reasoning fell from 54% to 0.2%
5. Inkling 2 Mini: a 20 GB open model that runs its middle layers three times
6. Signal blocks AI agents in iOS 27 and Android
7. Frostbite: AI designed a processor with a new instruction set [single source]
8. Trump threatens a tax on data centers of US companies abroad
9. A language model rewrites its own weights without gradients [single source]
10. Google employees against the cutoff of schools from Search

---

TOPIC 1

Alpöge and Claude announced a proof of the generalized Riemann hypothesis

SOURCES Levent Alpöge, 20.11 Alpo · Terence Tao, 20.11 Wordpress · Scott Aaronson, 20.11 Scottaaronson · BBC, 20.11 BBC · Simon Willison, 20.11 Simon Willison · @emollick, 20.11 @emollick · @TheRundownAI, 20.11 @TheRundownAI · HN Hacker News

Levent Alpöge is a mathematician at Anthropic. On 20.11 he published a four-page statement. He writes that he and Claude found a proof of the Riemann hypothesis, and the method also covers the generalized hypothesis (GRH). The Riemann hypothesis is open since 1859 and is one of the seven Millennium Prize Problems. The HN thread got 2,433 points and 1,642 comments, the most of the day.

Alpöge published only a formalization in Lean, a language for machine checks of proofs. There is no human text of the proof. The author explains this himself. The document with all details came to 250,000+ pages, and he did not want to publish such an "unreadable mass". Scott Aaronson counts 17.4 million lines of Lean code. BBC writes that Alpöge worked for about six weeks and used a non-public version of Claude. The timeline in the statement starts on 06.10 and ends with the item "16.11: GRH".

Terence Tao put this result last on his scale of how "digested" a result is. He advises readers to follow the check in the Zulip thread of the Lean community. A first analysis shows that the proof uses the full power of the modern Langlands program. Tao writes that the check will take time, but here at least a human author is available for talks.

Aaronson reminds readers of a weak point of Lean: kernel bugs in the past allowed a "proof" of 0=1. But he bets that the proof will stand.

Alpöge also writes about the time of the announcement: the day after the Anthropic stock market debut. He called this coincidence "terrible". He wrote that on Sunday he asked Dario Amodei for permission to announce. He says that the cynical explanation of the coincidence is false. On HN, an essay by Fernando Borretti reads the statement as a human surrender to the machine: "Claude in the center, we in the margins".

Why it matters: if the proof stands, this is the largest mathematical result with AI participation. But at the moment it is only a claim, and the community did not accept it yet. The Lean check tells only whether the code is correct. People must decide if the Lean statement really means GRH, and this will take time. The format itself is also new: a full proof without a human text.

TOPIC 2

On the same day: a human proof of GRH from Masullo, and people already look for holes in it

SOURCES Ravi Vakil, 20.11 Wordpress · GitHub, 20.11 GitHub · Terence Tao, 20.11 Wordpress · @amasad, 20.11 @amasad · HN Hacker News

A few hours after the Alpöge statement, mathematician Ravi Vakil published a "hurried announcement". Alessandro Maria Masullo has his own proof of GRH, and Vakil stresses: "It is all human, with no AI". Brian Conrad checked this proof for several months. The plan was to polish the text for a year and only then announce it.

Within 8 hours Vakil added an update. Several people already told him about problems in the proof. They found them with the help of different AI models. Vakil writes that the community must close all questions before it accepts the proof.

On the same day, Tao described two more results from the same group. Alpöge, Tristan Buckmaster and Matei Coiculescu, with "heavy AI assistance", proved blowup of solutions of the Euler equations in three dimensions. Alpöge and Buckmaster also showed finite-time blowup for the Navier-Stokes equations. Here there are papers for humans and a formalization in Lean. The Navier-Stokes paper has 79 pages, and the Lean library has about 600,000 lines.

Why it matters: in one day, two approaches to one problem appeared. The AI proof has a full machine check, but no human can read it. A human can read the human proof, but AI models already find problems in it. AI now plays on both sides: it writes proofs and it checks the proofs of others.

TOPIC 3

Market: Anthropic -9% on day two, Nvidia -14%, Nasdaq lost its 2026 growth

SOURCES The Guardian, 20.11 Guardian · NBC News, 20.11 Nbcnews · Reuters (MarketScreener), 20.11 Marketscreener · Financial Times, 20.11 FT · @KobeissiLetter, 20.11 @KobeissiLetter · HN Hacker News · HN Hacker News

Yesterday the digest wrote about the first day of trading: ANT closed 17% above the IPO price. On the second day the shares fell 9%, to $88.30, The Guardian writes. The offer price was $83, so almost none of the gain remains.

NBC reports that Nasdaq fell 5.3% in one day and lost all of its 2026 growth. The S&P 500 lost 3.7%, the Dow Jones 1.9%.

Both indexes are now in a correction, which means 10% below the peak. Nvidia fell about 14%. Asia had its worst week since 2020: Taiwan -6.7%, Kospi -5.5%, Nikkei -5%, SoftBank -14.5%.

A Reuters report made the Nvidia fall faster. Two sources of the agency say that the US Securities and Exchange Commission (SEC) examines the accounts of Nvidia. Reuters does not know what exactly the SEC examines or when it started. Yesterday the digest also wrote about a subpoena from the California attorney general. An IG analyst suggests that Anthropic "sucks out liquidity": investors sell other shares to buy ANT.

Why it matters: NBC gives the explanation of investors: slower data center construction, less money for AI and the Nvidia scandal. The largest IPO in history did not stop the trend. For companies that plan to raise money for AI, the window becomes narrower.

TOPIC 4

OpenAI on distillation: attack success on hidden reasoning fell from 54% to 0.2%

SOURCES OpenAI, 20.11 OpenAI · Semafor, 20.11 Semafor · @OpenAI, 20.11 @OpenAI · @emollick, 20.11 @emollick

Distillation is the training of one model on the answers of another model. OpenAI writes that this method can copy the abilities of a model without its safeguards. The company gives an example. Someone distills the cyber skills of a model and uses the copy for attacks that the original model would refuse.

In an internal study, OpenAI trained the same student model on two datasets. The student on exact reasoning traces of the teacher in general did better than the student on short reasoning summaries. Thus the company now gives special protection to hidden reasoning. An attack on the reasoning from a previous request had 54.0% success against GPT-6 Astra and 0.2% against GPT-6.2 Astra. These numbers come from an internal test of the company.

Each week OpenAI blocks from thousands to tens of thousands of accounts, and in summer the rate grew a lot. The company again names DeepSeek and Moonshot AI, the maker of

Kimi. Semafor adds a new OpenAI rule for resellers such as OpenRouter. Now the company controls each service between itself and the end user. Ethan Mollick suggests that the end of easy distillation can explain the slower progress of open models.

Why it matters: hidden reasoning becomes a trade secret, and companies protect it as they protect weights. For users of resellers, this means more checks. For open models, this can mean slower progress, but at the moment this is only the hypothesis of Mollick.

TOPIC 5

Inkling 2 Mini: a 20 GB open model that runs its middle layers three times

SOURCES Thinking Machines, 20.11 Thinkingmachines · Hugging Face, 20.11 Hugging Face · @mervenoyann, 20.11 @mervenoyann · HN Hacker News

Thinking Machines Lab released Inkling 2 Mini with open weights. It is a Mixture-of-Experts model with 26.4 billion parameters, and 3.4 billion of them are active. The main feature is a "loop": the model runs its middle layers three times for each token. Thus it reasons longer, but without extra tokens.

The company reports that the model leads among open models below 100 billion parameters: DeepSWE v1.1 48.2%, Vals Index v2.2 46.5%. The model takes about 20 GB of memory, so it runs on a laptop. On a plane without internet, researchers used it to build a flight simulator on a MacBook Pro. Hugging Face immediately added support in transformers, llama.cpp and MLX.

The company also checked whether the "loop" hides reasoning from oversight. The chain of reasoning became shorter, but stayed readable. On agentic sabotage tests the model did not become worse. The company admits a limit: the loop does not add knowledge, because facts depend on the number of parameters.

Why it matters: this is one of the first open examples of a loop architecture. Agentic coding at the level of small frontier models now works locally, without a cloud. For teams that must keep their code in-house, this is a practical option.

TOPIC 6

Signal blocks AI agents in iOS 27 and Android

SOURCES Signal, 20.11 Signal · @emollick, 20.11 @emollick · HN Hacker News

Meredith Whittaker, the president of Signal, announced new limits. Signal blocks screenshots in iOS 27 and on Android when it sees that an AI assistant tries to read the screen. Last year Signal did the same against Recall in Windows 11.

The user can turn off the protection in the settings, but then the system agent sees the chats of this user.

Signal writes that Gemini Intelligence and Onscreen Awareness in Siri "scan" the screen. Chat data then goes into the storage of the agent, and Signal no longer controls it. The company admits that the protection is incomplete: the agent can read files on the device. On macOS Signal has no protection at all, because Apple gives no tools for it.

Yesterday the digest mentioned that Apple builds its own agent for the Mac. On the same day, Ethan Mollick wrote that companies will annoy their customers when they ban the agents of these customers.

Why it matters: an OS-level agent sees all that a human sees, end-to-end encrypted chats included. Encryption protects the channel, and it does not protect the screen. App developers have no good method yet to tell an agent "no access here".

TOPIC 7

Frostbite: AI designed a processor with a new instruction set [single source]

SOURCES Architect Labs, 20.11 Architectlabs · HN Hacker News

The startup Architect Labs presented Frostbite, a general-purpose processor with a new architecture. The company writes that its AI systems did the full path alone: compiler, instruction set, microarchitecture, RTL, verification and the final GDS. GDS is the file that goes to the fab.

The base is interaction nets, a model of computation where each value has one owner and each step is local. Thus parallelism comes from the program itself, without threads and locks. On 25 tasks, 15 of them held back until the final evaluation, Frostbite gave 4.59 times better performance per watt. The baseline was a modern superscalar core. This core runs two instructions per cycle in program order.

The paper itself gives an important warning: the power numbers are "pre-silicon" at the moment. The chip does not exist yet, and all measurements come from a model of the company. There is no independent check.

Why it matters: AI already rewrites both code and the chip architecture itself. But the test here is the same as for any chip: real silicon. Until then, this is a claim of the company.

TOPIC 8

Trump threatens a tax on data centers of US companies abroad

SOURCES BBC, 20.11 BBC · Financial Times, 20.11 FT · The New York Times, 20.11 NYT

Donald Trump said that he will ban US companies from the construction of AI data centers abroad. He promised a "tremendous tax" to companies that build them. The explanation of Trump: "China will take all of them". He did not name a mechanism. He said that his team hears about the idea for the first time.

Meta and Google already have deals for data centers abroad. BBC reminds readers that on Monday the US introduced licenses for the purchase of advanced chips. Meanwhile, Canada offers itself as a site for data centers. Data centers were a topic of the 03.11 elections, and the Democrats promised stricter construction rules.

Why it matters: at the moment this is only a statement. But it hits plans that depend on data centers abroad:

Canada, the Middle East, Southeast Asia. Together with the chip licenses, the risk for such projects grew.

TOPIC 9

A language model rewrites its own weights without gradients [single source]

SOURCES arXiv, 16.11 arXiv · HN Hacker News

The preprint appeared on 16.11, and HN users brought it to the front page on 20.11. The authors gave an open 27-billion-parameter model its own weights, scores of its answers and the ability to run itself. The model found the cause of its errors and edited its weights itself, without gradients and without training.

The authors report that the self-edit made the model 6.6 points more accurate on new tasks. Over eight sessions the accuracy only grew. The model restored a damaged copy of itself to more than 80% of the initial accuracy. Smaller models of the same family cannot do this: they either edit nothing or break themselves.

Why it matters: the line between training and use of a model becomes less clear. The authors directly call this a step to recursive self-improvement. This is one preprint without peer review, and other teams must reproduce the result.

TOPIC 10

Google employees against the cutoff of schools from Search

SOURCES Mother Jones, 20.11 Motherjones · HN Hacker News

Two weeks ago Google announced a new rule. From next year, schools that turn off the Gemini AI tools will lose access to Google Search. Hundreds of employees signed a petition against this rule through the Alphabet Workers Union.

A quote from the petition: "Cutting off access to Search for schools that don't wish to use AI will not help persuade our users". Parents held protests in seven US cities. New York, Los Angeles and Chicago discuss other search engines.

The management received the letter, but did not answer the demands.

Why it matters: this is one of the first cases where access to a basic product depends on consent to AI. The reaction of schools will show how strong such a link is against organized

resistance.

---

misc

- **A society of AI scientists:** 936 agents from 35 labs worked for 12 weeks on a shared preprint server and wrote 936 studies. They publicly retracted errors, published negative results and founded their own peer-reviewed journal. A blind evaluation gave agents a higher score for rigor, and humans a higher score for novelty.
  Alphaxiv · @emollick
- **GPT-6.2 Astra on benchmarks:** XDOF reports that the model has 41.2% on the robotics benchmark Ember-0, against 19.6% for GPT-6 Astra in September. On Blueprint-Bench 2 from Andon Labs the model is first. Greg Brockman posted both results. @gdb · @gdb
- **MedChemBench:** OpenAI released a medicinal chemistry benchmark with 99 tasks from 117 papers and patents. GPT-6.2 Astra scored 84.5%, GPT-6 Astra 83.3%. OpenAI · @OpenAI
- **OpenSSL:** a PR closes a buffer overflow in `EVP_EncodeBlock` (CVE-2026-121481). The author writes that an attacker can reach the hole through `openssl s_client -proxy_pass`. OpenSSL did not publish an advisory yet. GitHub
- **GrapheneOS against Google:** the project writes that Google persuades manufacturers to drop support for GrapheneOS and for bootloader unlock. This is a claim of the project itself. Grapheneos
- **Mozilla and iOS:** almost three years after the DMA, no developer released an iOS browser without WebKit. Mozilla calls the Apple terms unworkable. Mozilla
- **When people stop writing code:** an essay by cekrem got 329 HN points. It says that without regular practice, the skill to review the code of others also becomes worse. Github
- **Applied Compute:** Palo Alto Networks buys the startup Applied Compute. The price is not public.
  @michaelzchen5
- **AI bots in replies:** @levelsio added a form where anyone can report an account with automatic AI replies.
  @levelsio
